AACR Abstract #876

A novel computational framework predicts synthetic lethal interactions between key regulators of DNA Damage Response and chromatin modifiers

Copies of this poster obtained through QR code are for personal use only and may not be reproduced without written permission of the authors

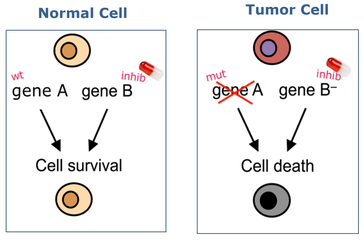
Anna M.L. Coenen-Stass[1], Magda Markowska[2], Magdalena Budzińska-Zaniewska[2], Krzysztof Kolmus[3], Ewa Szczurek[2] and Eike Staub[1]

[1]The healthcare business of Merck KGaA, Darmstadt, Germany
[2]Faculty of Mathematics, Informatics and Mechanics, University of Warsaw, Warsaw, Poland.
[3]Ardigen, Krakow, Poland

INTRODUCTION

The concept of synthetic lethality (SL) has made a pivotal impact for the development of the anti-cancer drug Olaparib, the first approved agent targeting the DNA Damage Response (DDR). Typically, SL is described as the interaction of two genes, whereby simultaneous inactivation of both genes results in cell death whereas loss of one gene can be tolerated. Given the importance of SL to develop highly selective anti-cancer therapeutics, large efforts have been made by the scientific community to identify these interactions, both experimentally and computationally.

Here, we present a novel computational framework harnessing large-scale cell line gene inactivation screens (DepMap, Project Score), as well as patient data from The Cancer Genome Atlas (TCGA), to discover known and novel SL gene pairs. Overall, we implemented six statistical tests considering gene dependency scores, genomic profiles, gene expression and patient survival as parameters. We further utilized data from public drug screening consortia to validate our top-ranking pairs. We applied this framework to a defined target space covering a set of genes relating to DDR, chromatin binding, cell cycle (overall > 1.4 Mio pairs were tested).



METHODS

To predict SL from cell line dependency screens or clinical data, 5 statistical tests are performed. The results of these tests are integrated using rank aggregation methods. Finally, public and internal drug screening data is exploited to provide validation evidence for the potential SL gene pairs.

SL prediction: gene dependency

Data preparation

Gene dependency datasets from the DepMap portal:
- CRISPR (DepMap 22Q1 Public, Chronos)
- CRISPR (Project Score, Chronos)
- RNAi (Achilles+DRIVE+Marcotte, DEMETER2)

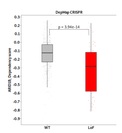
Cell Lines	Genes
	001001
	110011
	001100
	100001

Gene expression, mutation and copy number data was obtained from DepMap and cell model passport Portal. The genomic read-outs were converted into a binary format to indicated a loss of function (LoF) mutation or deletions. To determine LoF mutations, only variants with a predicted high impact or highly probable deleterious mismatch mutations were considered.

Synthetic partner inactivation dependency (SPID)

Hypothesis: Cell lines with gene A (ARID1A) mutation should have a higher sensitivity for gene B (ARID1B) inactivation if a SL interaction exists.

A one-sided Wilcoxon rank-sum test is performed, and the Cohen's D effect size is calculated.



Synthetic partner enrichment analysis (SPEA)

Hypothesis: The cell lines most sensitive to Gene B (ARID1B) deletion are also enriched for Gene A (ARID1A) LoF mutations.

An enrichment analysis is performed, based on the known gene set enrichment analysis (GSEA) algorithm which uses a permutation-based test and a Kolmogorow-Smirnoff statistic to compute the significance of enrichment.

SL prediction: Clinical data

Data preparation

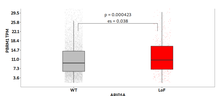
The following TCGA datasets were obtained from the Xena data portal or from GDC using the TCGAbiolinks R package:
- Gene expression
- Mutations
- Copy Number Variations
- Survival

Binary matrixes for mutation and copy number alterations were generated as described in the cell line section.

Survival of the fittest (SoF)

Hypothesis: There is a compensatory effect for loss of gene A resulting in an increased expression of gene B.

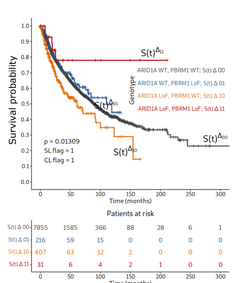
Test for avoidance of simultaneous deletion of a gene and low expression of its partner with a one-sided Wilcoxon rank-sum test. Jerby-Arnon, Cell 2014; Szczurek, Int J Cancer 2013. Cohen's D effect size.



survLRT

Hypothesis: Loss of both genes constituting a synthetic lethal pair will increase patient survival and decrease tumor fitness.

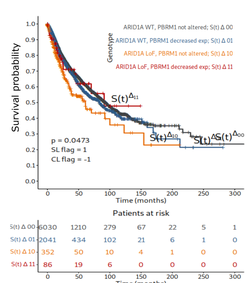
The survLRT method (Matlak and Szczurek 2017) is likelihood ratio test used to estimate the tumor fitness with a given genotype g from survival data of patients.



iterative survLRT

Hypothesis: Loss of both genes constituting a synthetic lethal pair will increase patient survival and decrease tumor fitness.

iSurvLRT method is an adaption to survLRT utilizing gene expression of gene B instead of mutation. The ideal threshold is identified iteratively.



Ranking

- The *RobustRankAggreg* R package (Kolde et al.) was utilised to integrate ranked gene pair lists using both p-values information and effect sizes where applicable (SPID, SPEA, SoF).
- The geomean method was chosen as best performing rank aggregation method.
- Cell line and patient data was aggregated separately and combined after an assessment of list overlapping with the *orderedList* package (Yang et al.).

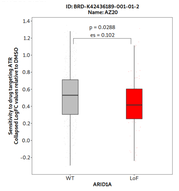
SL gene pair validation

Data preparation

Results from the following drug cell line screens were obtained from the DepMap portal:
- Drug sensitivity IC50 (Sanger GDSC1)
- Drug sensitivity IC50 (Sanger GDSC2
- Drug sensitivity (PRISM Repurposing Primary Screen) 19Q4)

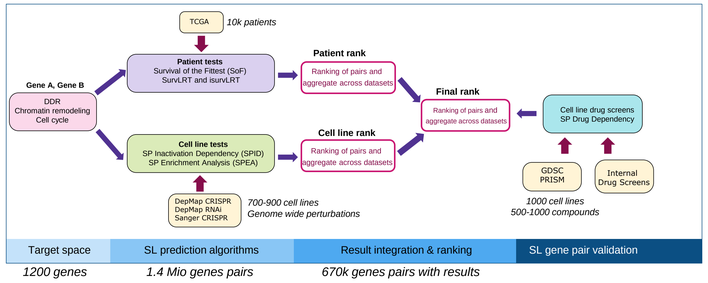
Hypothesis: Cell lines with a Gene A inactivation should be more susceptible to a drug targeting gene B

One-sided Wilcoxon rank sum test. Cohen's D effect size.



RESULTS

SL prediction framework and target space



Can we rediscover known SL pairs?

Examples of top-ranking pairs with previous SL evidence:

- TP53 and CHK2
- ARID1A and ARID1B
- TP53 and TP53BP1
- TP53 and ATM
- SMARCA4 and SMARCA2
- RB1 and SKP2

Potential SL partners of clinical DDRi are enriched for chromatin binders

- Gene ontology analysis of the top 20 ranking predicted SL partners for ATR, ATM and DNAPK
- 5 genes were ranked in the top 20 for all three DDR kinases (TP53 , CREBBP, KMT2D, EP300, PDS5B)
- High ranking genes are enriched for histone (de) methyltransferases, histone (de)acetylation and SWI/SNF complex members

Molecular Function

Name	GO ID	p-value	q-value FDR B&H
chromatin binding	GO:0003682	3,63E-16	1,20E-13
histone-lysine N-methyltransferase activity	GO:0018024	3,80E-11	6,28E-09
histone methyltransferase activity	GO:0042054	2,37E-10	2,61E-08
protein-lysine N-methyltransferase activity	GO:0016279	5,50E-10	3,47E-08
lysine N-methyltransferase activity	GO:0016278	6,27E-10	3,47E-08

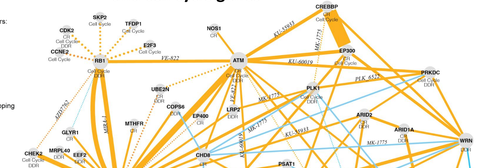
Biological Process

Name	GO ID	p-value	q-value FDR B&H
chromosome organization	GO:0051276	8,42E-34	2,43E-30
chromatin organization	GO:0006325	1,46E-28	2,11E-25
histone modification	GO:0016570	8,27E-21	7,94E-18
covalent chromatin modification	GO:0016569	1,48E-20	1,07E-17
peptidyl-amino acid modification	GO:0018193	1,52E-15	8,75E-13

Cellular Component

Name	GO ID	p-value	q-value FDR B&H
chromosome	GO:0005694	2,46E-16	6,40E-14
nuclear protein-containing complex	GO:0140513	2,10E-13	2,73E-11
catalytic complex	GO:1902494	2,42E-08	2,10E-06
chromatin	GO:0000785	9,54E-08	6,20E-06
SWI/SNF superfamily-type complex	GO:0070603	1,46E-07	7,57E-06

- Venn diagrams indicating the overlap between the top 20 ranking SL partners and top 20 sensitizing mutations in public drug screening data.



Example 1: The KMT2 family

- KMTs: Lysine-specific histone methyltransferases
- Histone modifications modulate chromatin accessibility and are required for DNA repair and genome stability

SL prediction algorithms | SL gene pair validation

- Cell lines with KMT2D LoF mut are more sensitive to CRISPR/ RNAi mediated inactivation of ATM
- Stomach cancer patients with KMT2D and ATR LoF mutations survive longer, and the tumor fitness is decreased.
- Patients with KMT2C and ATM LoF mutation have better survival (data not shown).

KMT2D LoF mut also increases sensitivity for other ATRi (VE81, AZD6738, AZ20) and ATMi (KU-55933, KU−60019, CP46672) in the GDSC screen.

KMT2 & Cancer

- KMT2D/C are highly mutated in bladder cancer but also others (lung squamous, uterine, H&N, gastric)
- Several papers indicate a connection of KMT2's to DDR gene regulation and HRD (Rampias et al., Chang et al.)

Example 2: The CCCTC-binding factor (CTCF)

- Zinc finger, suggested as key regulator for genome 3D structure and stability.
- Functional interaction with cohesion complex likely, mediates long range genomic interactions

SL prediction algorithms | SL gene pair validation

- Cell lines with CTCF LoF mut are more sensitive to CRISPR/ RNAi mediated inactivation of ATM
- Patients with CTCF LoF and decreased ATM expression have better survival.
- CTCF LoF mut also increases sensitivity for ATRi (VE-821, VE-822, AZD6738 in the GDSC screen

CTCF & Cancer

- CTCF is recruited to DSBs, may be involved in confining γH2AX foci to DSB sites via BRCA2 (Tanwar et al.)
- CTCF genetic alterations in endometrial carcinoma are pro-tumorigenic: role for CTCF in the regulation of cellular polarity of endometrial glandular epithelium (Marshall et al.)
- High frequency strand slippage mutations in CTCF in MSI-positive endometrial cancers: suggested CTCF acting as a haploinsufficient tumor suppressor. (Zighelboim et al).

CONCLUSION

- A modular framework to predict SL gene pairs was developed and applied to defined target space.
- Top ranking SL partners of the key DDR regulators ATM, ATR and DNA-PK were highly enriched for chromatin modifiers, i.e., histone (de)methylases, histone (de)acetyltransferases and members of SWI/SNF family.
- The results provide new biomarker hypotheses for further validation and suggest that cancers with a high mutation rate in chromatin modifying genes may be efficiently targeted by DDRi.

Summary of SL evidence network for DDR, chromatin modifier and cell cycle genes

Disclosures: Anna Coenen-Stass and Eike Staub are employees of the healthcare business of Merck KGaA, Darmstadt, Germany. Krzysztof Kolmus is an employee of Ardigen, Krawkow, Poland. Ewa Szczurek's group has received research funding by the healthcare business of Merck KGaA, Darmstadt, Germany. This research was supported by the healthcare business of Merck KGaA, Darmstadt, Germany (CrossRef Funder ID: 10.13039/100009945)

Abstract #876 Presented at the American Association of Cancer Research (AACR) Annual Meeting 2022| April 8-13| New Orleans, USA

Contact: anna.coenen-stass@emdgroup.com